BU-820-M

EXPLICIT FORMULAS FOR CONFIDENCE INTERVAL ESTIMATION IN DISCRETE DISTRIBUTIONS

by

June 1983
Revised May 1983

Charles E. McCulloch and George Casella*

Biometrics Unit, Cornell University, Ithaca, New York 14853

Explicit formulas for confidence intervals for the parameters of the binomial, Poisson and negative binomial distributions are presented which require only F and χ^2 tables. The derivation is at a level suitable for a first mathematical statistics course. Properties of the derived intervals are discussed.

KEY WORDS: Binomial; Poisson; Negative binomial; Coverage probability.

*Charles E. McCulloch is Assistant Professor and George Casella is Associate Professor, both in the Biometrics Unit, Cornell University, Ithaca, NY 14853. Dr. Casella's research was supported by National Science Foundation Grant No. MCS81-02541. This is paper #BU-820-M in the Biometrics Unit series.

1. INTRODUCTION

For the most commonly used discrete distributions (binomial, Poisson, and negative binomial), explicit formulas exist for confidence intervals that require only standard statistical tables (F and χ^2). The binomial and Poisson formulas are available in some places but where they are available the properties of the intervals are not clearly stated or are misstated. In many other cases, the formulas are apparently unknown. This article has several objectives. The first objective is to present a unified approach to the construction of the intervals at a mathematical level accessible to a first course in mathematical statistics. Another objective is to call attention to these formulas, for they sometimes prove useful in small-sample applications. Also, in some instances, intervals derived from large-sample considerations are not adequate and these explicit formulas or some modification of the large-sample intervals should be used. Finally, we use the explicit formulas to derive some properties of the intervals.

The formulas we present are not all new. The binomial interval is an explicit formula based on the F distribution, and produces the intervals first tabled by Clopper and Pearson (1934). For example, these formulas can be found in the introductory book by Neter, Wasserman and Whitmore (1982). The earliest references we can find to the explicit formulas are Hald (1952) and Brownlee (1960), but earlier papers, such as Stevens (1950), derive the intervals in terms of incomplete beta integrals.

Despite this, the explicit formulas for the binomial intervals often seem to be unknown. Ghosh (1979), in a recent article, argues for the need for good approximate binomial intervals. When discussing the

Clopper-Pearson intervals (p', p"), he states, "These intervals require extensive tables of p' and p", and this aspect may be regarded as an obstacle... ." The article recommends an approximate interval which is reasonably complicated to calculate. Yet, the Clopper-Pearson intervals can, in many cases, be easily calculated using only F-tables, such as those found in Snedecor and Cochran (1980). Granted, they may be less intuitive than the approximate interval

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} \quad , \tag{1.1}$$

but this interval has poor coverage properties (Ghosh, 1979; Blyth and Still, 1983). The performance of (1.1) is greatly improved with the addition of a continuity correction, but the interval then loses its intuitive appeal.

In the standard mathematical statistics text by Mood, Graybill and Boes (1974), the implicit formulas for the Clopper-Pearson intervals, in terms of binomial sums, are listed, but no mention is made of an explicit solution. A derivation of the result is not beyond the theoretical level of the text. As Ghosh (1979) mentions, many elementary or applied texts list only (1.1) as an approximate interval for p. Often this is done without mention of the performance of the interval for small n, or p near 0 or 1. It would not be difficult to also list the Clopper-Pearson formulas as Neter, Wasserman and Whitmore (1982) have done. Kendall and Stuart (1979) list several charts and tables and mention the use of the incomplete beta function, but not the F distribution.

To get some idea whether these formulas are known, and are employed by professional statisticians, we conducted an informal survey of our colleagues. They were asked what methods they would use to construct a $1-\alpha$ confidence interval for a binomial parameter (for both large and small samples) and what what methods they would use to construct $1-\alpha$ intervals for a Poisson parameter. The results are summarized in Table 1.

Table 1: Results of Informal Survey

	Method			
	Tables	Statistical Method	Normal Approx.	Explicit Formula
Binomial	53.3%	20.0%	20.0%	6.7%
Poisson	20.0%	13.3%	60.0%	6.7%

Of the people who indicated that they would use the Statistical Method, none mentioned that the binomial and Poisson sums could be converted to recognizable integrals. We feel that the results of our survey are strongly indicative of the fact that many of these simple confidence techniques are not well known among statisticians, especially when we note that the 6.7% aware of these techniques represents one person who has done research in this area.

We see several advantages to the explicit formulas for confidence intervals for the discrete distributions. In a mathematical statistics course the formulas can be derived to illustrate what Mood, Graybill and Boes (1974) call the Statistical Method of deriving confidence intervals. In a mathematical statistics course or a statistical methods course they offer a demonstration of large sample approximations. A common question in statistics courses is "How big does n have to be before the central limit theorem takes effect?". The approximate intervals found by (1.1) can be compared to the derived intervals to investigate the quality of the large sample approximation. In addition to a comparison, the derived intervals offer a fall-back position for the cases in which the approximation may be dubious.

The formulas can also be useful in practice for similar reasons. For small sample sizes they are computationally as simple as the recommended approximations in Ghosh (1979) and Blyth and Still (1983). For larger sample sizes, however, calculation of the intervals presented here can become somewhat laborious. This will occur when the required cut-off values cannot be read directly from a table, and interpolation becomes necessary. This extra work may be justified if the normal approximation is suspect but, generally speaking, for large samples the normal approximation is reasonably accurate and requires fewer calculations.

The Poisson formulas are presented in an early reference (Garwood, 1936), where the original derivation is fiducial. The derivation we present in Section 2, although algebraically equivalent, is more appropriate for present-day mathematical statistics courses. For the negative binomial, we do not know of previous confidence interval estimates in implicit or explicit form. Clemans (1959) does give graphical intervals for the geometric distribution, however.

The explicit formulas are useful in doing statistical research. Having an explicit representation for the endpoints of the intervals often makes theoretical manipulations easier.

Even when the explicit formulas are known, the properties of the intervals are not well understood. For example, the two-sided intervals for the binomial are described as exact in Neter, Wasserman and Whitmore (1982) and Miettinen (1970). In Johnson and Kotz (1969) they are described as approximate (with italics for emphasis). Mood, Graybill and Boes (1974) require, in their definition of a confidence interval, that the coverage probability be independent of the parameter. One is left with the implication that finding an interval by their method would guarantee this, which cannot be true using nonrandomized intervals with a discrete distribution.

In order to make our presentation clear, and to clear up past ambiguities about the properties of these intervals, we require some terminology. Let $CP(\theta,\alpha)$ denote the coverage probability of a nominal $1-\alpha$ confidence interval for a parameter $\theta \in \Theta$. In practice, the nominal level (the value used to determine the appropriate tabled cutoffs) will be the quoted level of confidence. However, in many cases, this nominal level will not equal the infimum of the coverage probabilities. For example, when illustrating the use of the explicit formulas for the binomial distribution, Brownlee (1960) calls the intervals 95% confidence intervals, which is a misnomer. We can classify a confidence procedure, using its coverage probability, according to the following definitions.

<u>Definition 1.1</u>: A confidence procedure with coverage probability $CP(\theta,\alpha)$ is called

(i) <u>sharp</u> if $\inf_{\theta\in\Theta} CP(\theta,\alpha) = 1-\alpha$

(ii) <u>exact</u> if $CP(\theta,\alpha) = 1-\alpha$ for all $\theta\in\Theta$

(iii) <u>conservative</u> if $\inf_{\theta\in\Theta} CP(\theta,\alpha) > 1-\alpha$

(iv) <u>approximate</u> if $CP(\theta,\alpha) \approx 1-\alpha$ for all $\theta\in\Theta$.

It will be seen that, for the binomial distribution, the intervals derived are either conservative or sharp, with the former case the more likely. However, surprisingly, for the Poisson and negative binomial, the resulting intervals are always sharp.

2. Derivation of the Intervals

2.1. General Theory

Before treating the special cases of the binomial, Poisson and negative binomial, we outline the methodology behind the construction of the confidence procedure. Some of the original derivations of these intervals were fiducial in nature so we present our derivations within the more common Neyman-Pearson framework of confidence intervals. Let X be a discrete random variable with distribution dependent on $\theta\varepsilon\Theta$. Given α, define for every x,

$$\overset{1}{S}_{\alpha/2}(x) = \{\theta\varepsilon\Theta : P_\theta(X \le x) \ge \alpha/2\}$$

and (2.1)

$$\overset{2}{S}_{\alpha/2}(x) = \{\theta\varepsilon\Theta : P_\theta(X \ge x) \ge \alpha/2\} \quad .$$

We can then state the following theorem.

<u>Theorem 2.1</u>: The coverage probability of the random set

$$S_\alpha(X) = \overset{1}{S}_{\alpha/2}(X) \cap \overset{2}{S}_{\alpha/2}(X) \tag{2.2}$$

is at least $1-\alpha$.

<u>Proof</u>: We will show that each of the random sets $\overset{1}{S}_{\alpha/2}(X)$ and $\overset{2}{S}_{\alpha/2}(X)$ satisfies

$$P_\theta\left(\theta\varepsilon \overset{i}{S}_{\alpha/2}(X)\right) \ge 1 - \frac{\alpha}{2} \quad \text{for all } \theta, \quad i=1,2 \ . \tag{2.3}$$

From (2.3) it then follows that

$$P_\theta\left(\theta\varepsilon S_\alpha(X)\right) = P_\theta\left(\theta\varepsilon S^1_{\alpha/2}(X)\right)+P_\theta\left(\theta\varepsilon S^2_{\alpha/2}(X)\right)-P_\theta\left(\theta\varepsilon S^1_{\alpha/2}(X) \cup S^2_{\alpha/2}(X)\right)$$

$$\begin{aligned} &\geq \left(1 - \frac{\alpha}{2}\right)+\left(1 - \frac{\alpha}{2}\right) - P_\theta\left(\theta\varepsilon S^1_{\alpha/2}(X) \cup S^2_{\alpha/2}(X)\right) \\ &= 1-\alpha+\left[1-P_\theta\left(\theta\varepsilon S^1_{\alpha/2}(X) \cup S^2_{\alpha/2}(X)\right)\right] \\ &\geq 1-\alpha \qquad \text{for all } \theta \quad , \end{aligned} \tag{2.4}$$

and the theorem will be established.

Consider the set $S^1_{\alpha/2}(X)$. For every θ, define a sample point $x_{\alpha/2}(\theta)$ by

$$P_\theta\left(X < x_{\alpha/2}(\theta)\right) \leq \alpha/2 \quad \text{and} \quad P_\theta\left(X \leq x_{\alpha/2}(\theta)\right) \geq \alpha/2 \quad . \tag{2.5}$$

We then have for every x

$$\begin{aligned} \theta\varepsilon S^1_{\alpha/2}(x) &\text{ if and only if } P_\theta(X \leq x) \geq \alpha/2 \quad , \\ &\text{ if and only if } x \geq x_{\alpha/2}(\theta) \end{aligned} \tag{2.6}$$

and thus

$$P_\theta\left(\theta\varepsilon S^1_{\alpha/2}(X)\right) = P_\theta\left(X \geq x_{\alpha/2}(\theta)\right) \geq 1 - \frac{\alpha}{2}$$

from (2.5). A similar argument will establish that $P_\theta\left(\theta\varepsilon S^2_{\alpha/2}(X)\right)\geq 1 - \frac{\alpha}{2}$, proving the theorem. ‖

In general, we are interested in confidence intervals rather than confidence regions. It is easy to see that the region $S_\alpha(x)$ will be an interval if both $P_\theta(X \geq x)$ and $P_\theta(X \leq x)$ are monotone functions of θ for every x. If, for example, $P_\theta(X \geq x)$ is increasing in θ (which is the case for many common distributions), then

$$\begin{aligned} &P_\theta(X \geq x) \geq \alpha/2 \text{ if and only if } \theta \geq \theta_L(x) \\ \text{and} \quad &P_\theta(X \leq x) \geq \alpha/2 \text{ if and only if } \theta \leq \theta_U(x) \end{aligned} \tag{2.7}$$

and the intervals $\left(\theta_L(X),\theta_U(X)\right)$ then form a set of $1-\alpha$ confidence intervals.

We also note that the set of intervals $(\theta_L(X),\theta_{max})$ are a set of sharp $1-(\alpha/2)$ confidence intervals. (In general, each of $S^1_{\alpha/2}(X)$ and $S^2_{\alpha/2}(X)$ are sharp $1-(\alpha/2)$ confidence regions.) This shows that the two-sided intervals formed by this method are usually conservative, since they are essentially constructed by overlapping two one-sided intervals and employing the inequality $P(A)+P(B)-1 \le P(A\cap B)$, the simplest version of the Bonferroni inequality (Miller, 1977).

Although, in this development, α was divided equally between S^1 and S^2, in general it can be divided in any way and the above construction will still yield a $1-\alpha$ confidence region. Dividing α equally, however, has some advantages. Firstly, if the Central Limit Theorem is applicable then, asymptotically, intervals with equal tail probabilities are optimal, since they are optimal for the normal density. Secondly (for commonly chosen α-levels), dividing α equally will usually facilitate table look-ups.

We also note that Theorem 2.1 is similar to theorems which relate α-level hypothesis tests to $1-\alpha$ confidence intervals. For example, one can see that, for fixed $\theta=\theta_0$, the set $T^1_{\alpha/2}(\theta_0) = \{x:P_{\theta_0}(X\le x)\ge\alpha/2\}$ is an acceptance region for testing the null hypothesis $H_0:\theta=\theta_0$ vs. $H_A:\theta>\theta_0$.

2.2 The Binomial Distribution

One of the main points of this article is to illustrate that, for many common distributions, $P_\theta(X\ge x) = \alpha/2$ can be solved for explicitly in terms of commonly available tabled values of continuous distributions. If $X \sim \text{binomial}(n,\theta)$, i.e., $P_\theta(X=x) = \binom{n}{x}\theta^k(1-\theta)^{n-x}$, then we can write

$$P_\theta(X \geq x) = \sum_{k=x}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k}$$
$$= \sum_{k=x}^{n} \binom{n}{k} \int_0^\theta \left[\frac{d}{dt}\, t^k (1-t)^{n-k}\right] dt \quad . \tag{2.8}$$

Now $(d/dt)\left(t^k(1-t)^{n-k}\right) = \left(kt^{k-1}(1-t)^{n-k} - (n-k)t^k(1-t)^{n-k-1}\right)$, and interchanging the order of summation and integration we have

$$P_\theta(X \geq x) = \int_0^\theta \left\{ \sum_{k=x}^{n} \binom{n}{k}\left[kt^{k-1}(1-t)^{n-k} - (n-k)t^k(1-t)^{n-k-1}\right]\right\} dt \quad . \tag{2.9}$$

The first sum in braces can be rewritten as

$$\sum_{k=x}^{n} \binom{n}{k} kt^{k-1}(1-t)^{n-k} = \sum_{j=x-1}^{n-1} \binom{n}{j+1}(j+1)t^j(1-t)^{n-j-1} \qquad (j=k-1)$$
$$= \sum_{j=x-1}^{n-1} \binom{n}{j}(n-j)t^j(1-t)^{n-j-1} \quad . \tag{2.10}$$

Now substituting back into (2.9), and canceling the common terms, we have

$$P_\theta(X \geq x) = \int_0^\theta \left\{\binom{n}{x-1}(n-x+1)t^{x-1}(1-t)^{n-x}\right\} dt$$
$$= \frac{\Gamma(n+1)}{\Gamma(x)\Gamma(n-x+1)} \int_0^\theta t^{x-1}(1-t)^{n-x} dt \quad , \tag{2.11}$$

where the last integral can now be recognized as a beta probability.

Thus we have

$$P_\theta(X \geq x) = P(U < \theta) \tag{2.12}$$

where U follows a beta distribution with parameters x and n-x+1. Next, recall that $\left(\frac{n-x+1}{x}\right)\left(\frac{U}{1-U}\right)$ is distributed as an F random variable with 2x and 2(n-x+1) degrees of freedom. Hence,

$$P_\theta(X\geq x) = P(U<\theta) = P\left[\frac{n-x+1}{x}\left(\frac{U}{1-U}\right) < \frac{n-x+1}{x}\left(\frac{\theta}{1-\theta}\right)\right]$$

$$= P\left(F_{2x,2(n-x+1)} < \frac{n-x+1}{x}\,\frac{\theta}{1-\theta}\right) \tag{2.13}$$

$$= P\left(F_{2(n-x+1),2x} > \frac{x}{n-x+1}\,\frac{1-\theta}{\theta}\right)$$

since $F_{2x,2(n-x+1)} \sim 1/F_{2(n-x+1),2x}$. So, if we choose $\theta(x)$ to satisfy

$$\frac{x}{n-x+1}\,\frac{1-\theta(x)}{\theta(x)} = F_{2(n-x+1),2x,\alpha/2} \quad , \tag{2.14}$$

where, in general, $F_{\nu_1,\nu_2,\gamma}$ satisfies $P\left(F_{\nu_1,\nu_2} > F_{\nu_1,\nu_2,\gamma}\right) = \gamma$, we have $P_{\theta(x)}(X\geq x) = \alpha/2$. Moreover, from the monotonicity of $P_\theta(X\geq x)$, we have $P_\theta(X\geq x) \leq \alpha/2$ for $\theta < \theta(x)$. We can thus define the lower bound for our confidence interval as the solution to (2.14), i.e.,

$$\theta_L(x) = \frac{1}{1 + \frac{n-x+1}{x} F_{2(n-x+1),2x,\alpha/2}} \quad . \tag{2.15}$$

Similarly, it can be shown that $P\ (X\leq x) = P(V>\theta)$, where $V \sim$ Beta(n-x,x+1). A similar argument with the F-distribution then yields the upper bound

$$\theta_U(x) = \frac{\frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}}{1 + \frac{x+1}{n-x} F_{2(x+1),2(n-x),\alpha/2}} \quad . \tag{2.16}$$

When $x=0$ or n, these formulas give zero degrees of freedom and division by zero for one of the endpoints. By using (2.5) directly these can be resolved as $\theta_L(0) = 0$ and $\theta_U(n) = 1$. Thus, for all θ, the probability that the random interval $\left(\theta_L(X),\ \theta_U(X)\right)$ contains θ is at least $1-\alpha$.

The use of these intervals in practice will, of course, be limited to those cases for which appropriate F cutoff values are readily obtainable. Access to an F-table such as that found in Snedecor and Cochran (1980) (which is a reproduction of the *Biometrika* table) will generally allow easy construction of these intervals for small values of n. In particular, for $n \leq 10$ and $\alpha = .01$, .05 or .1, these intervals can be used with little or no interpolation. For $\alpha = .1$ (90% confidence) the tables are quite extensive even for large n, and should not require any calculations more complicated than linear interpolation.

As an example, if $n = 12$ and $x = 7$, the relevant F values are $F_{16,10,.05} = 2.82$ and $F_{12,14,.05} = 2.53$. Then,

$$\theta_L = \frac{1}{1 + \frac{6}{7}(2.53)} = .316, \qquad \theta_U = \frac{\frac{8}{5}(2.82)}{1 + \frac{8}{5}(2.82)} = .819 \quad .$$

2.3. The Poisson and Negative Binomial Distributions

The method illustrated above can also be used to calculate confidence intervals for the Poisson and negative binomial distributions. We will only sketch the derivations here.

If $X \sim \text{Poisson}(\theta)$, i.e., $P_\theta(X=x) = e^{-\theta}\theta^x/x!$, then

$$\begin{aligned}
P_\theta(X \geq x) &= \sum_{k=x}^{\infty} e^{-\theta}\theta^k/k! = \sum_{k=x}^{\infty} \frac{1}{k!}\int_0^\theta \left[\frac{d}{dt}(e^{-t}t^k)\right] dt \\
&= \sum_{k=x}^{\infty} \frac{1}{k!} \int_0^\theta \left(kt^{k-1}e^{-t} - t^k e^{-t}\right) dt \qquad (2.17)\\
&= \frac{1}{(x-1)!} \int_0^\theta t^{x-1}e^{-t}dt \\
&= P(\chi^2_{2x} < 2\theta)
\end{aligned}$$

where χ^2_{2x} denotes a chi-squared random variable with 2x degrees of freedom. Thus, we find that a 1-α confidence interval for θ is given by $\left(\theta_L(X),\ \theta_U(X)\right)$ where

$$\theta_L(x) = \tfrac{1}{2}\chi^2_{2x,1-\alpha/2} \qquad \theta_U(x) = \tfrac{1}{2}\chi^2_{2(x+1),\alpha/2}\ , \tag{2.18}$$

and $\chi^2_{\nu,\alpha}$ satisfies $P(\chi^2_\nu > \chi^2_{\nu,\alpha}) = \alpha$. If $X_1, X_2, \ldots, X_n$ are iid Poisson(θ), then $Y = \Sigma X_i \sim$ Poisson($n\theta$), and it follows that

$$\theta_L(Y) = \frac{1}{2n}\chi^2_{2Y,1-\alpha/2} \qquad \theta_U(Y) = \frac{1}{2n}\chi^2_{2(Y+1),\alpha/2} \tag{2.19}$$

is a confidence interval for θ.

The negative binomial distribution can easily be treated by exploiting its relationship to the binomial. If $X \sim NB(r,\theta)$, i.e.,

$$P_\theta(X=x) = \binom{r+x-1}{x}\theta^r(1-\theta)^x, \qquad x = 0,1,2,\ldots$$

then $P_\theta(X \geq x) = P_\theta(W < r)$, where $W \sim$ Binomial($x+r-1,\theta$). Thus, the binomial derivation can be applied to yield a confidence interval given by

$$\theta_U(x) = \frac{\frac{r}{X}F_{2r,2X,\alpha/2}}{1 + \frac{r}{X}F_{2r,2X,\alpha/2}}$$

$$\theta_L(x) = \frac{1}{1 + \frac{X+1}{r}F_{2(X+1),2r,\alpha/2}} \tag{2.20}$$

If $X_1, X_2, \ldots, X_n$ are independent $NB(r_i,\theta)$ then $\sum X_i \sim NB\left(\sum_{i=1}^n r_i,\theta\right)$ and the same formula can be used with $r = \sum_{i=1}^n r_i$.

2.4. Some Remarks on the Derivations

The derivations of these confidence intervals are an application of the Fundamental Theorem of Calculus to the terms of $P_\theta(X \geq x)$. For example, for the binomial we write

$$P_\theta(X=x) = \binom{n}{k} \theta^k(1-\theta)^{n-k} = \int_0^\theta \frac{d}{dt} \binom{n}{k} t^k(1-t)^{n-k}dt$$
$$= \int_0^\theta \binom{n}{k} \left[kt^{k-1}(1-t)^{n-k} - t^k(n-k)^{n-k-1}\right] dt \quad .$$

We then use the above representation in

$$P_\theta(X \geq x) = \sum_{k=x}^{n} P_\theta(X=k) \quad .$$

This technique will apply, in general, to any distribution for which $P_\theta(X \geq x)$ is a differentiable function of θ. The resulting integral may not be easily recognized, as was the case here, but it will still provide, in many cases, an easier way to obtain a set of $1-\alpha$ intervals.

The relationships used in deriving these intervals, particularly that of the binomial and beta distributions, are quite well known. Furthermore, there are many alternative, and perhaps more elegant, methods for deriving equations such as (2.12). However, the method described in this subsection, using the Fundamental Theorem of Calculus, seems to have two distinct advantages. One, the same technique is applicable to a variety of distributions and two, it is based only on elementary mathematical and statistical concepts.

3. Properties of the Intervals

In this section we describe some properties of the confidence intervals derived for the binomial, Poisson, and negative binomial distribution. In Section 3.1 we list some generally desirable properties of confidence regions, and show that the intervals described here have these properties. In particular, it will be seen that verification of these properties is made extremely easy by the fact that we have explicit formulas for the endpoints. In Section 3.2 we further investigate coverage probabilities. Contrary to what seems to have become common belief, the binomial intervals are, in general, conservative. Only for some specific values of α are these intervals sharp. The Poisson and negative binomial intervals are sharp for all α.

3.1. General Properties

Blyth and Still (1983) list several properties that are desirable for binomial confidence regions to possess. Three of them are applicable more generally:

(i) *Interval-valued.* The realized confidence region should always be an interval.

(ii) *Monotone in α.* For fixed x and n, the lower endpoint should increase in α, and the upper endpoint should decrease in α.

(iii) *Monotone in x.* For fixed n, if $E_\theta[X]$ is increasing (decreasing) in θ, the interval endpoints (both upper and lower) should be increasing (decreasing) in x.

We have already seen that, as long as the distribution function is monotone in the tails, our confidence regions will always be intervals. This is so for the binomial, Poisson, and negative binomial.

Because of the explicit formulas available for the endpoints, and the many well-known properties of the χ^2 and F distributions, these properties are relatively easy to check. Consider first the Poisson intervals, where

$$\theta_L(x) = \tfrac{1}{2}\chi^2_{2x,1-\alpha/2}, \quad \theta_U(x) = \tfrac{1}{2}\chi^2_{2(x+1),\alpha/2} \quad . \tag{3.1}$$

Monotonicity in x follows immediately from the fact that the χ^2 distribution is stochastically increasing in its degrees of freedom. That is, if $p_1 > p_2$, then $P(\chi^2_{p_1} > a) > P(\chi^2_{p_2} > a)$, hence $\chi^2_{\nu,\alpha}$ is increasing in ν. Similarly, it again follows quickly from (3.1) and the properties of χ^2 that the intervals are monotone in α.

For the binomial and negative binomial intervals, we must deal with the F-distribution. But again, the properties of this distribution, together with the explicit formulas, make checking these conditions quite easy. For example, consider the lower binomial bound

$$\theta_L(x) = \frac{1}{1 + \frac{n-x+1}{x} F_{2(n-x+1),2x,\alpha/2}} \quad . \tag{3.2}$$

From the properties of the F-distribution, it is easy to see that $\theta_L(x)$ is increasing in α. Also, writing $[(n-x+1)/x]F_{2(n-x+1),2x} = \chi^2_{2(n-x+1)}/\chi^2_{2x}$, and using the fact that a χ^2-variable has monotone likelihood ratio in its degrees of freedom, it can be established that $\theta_L(x)$ is increasing in x.

Therefore, for the three distributions considered here and, in general, for reasonably behaved distributions, the confidence regions derived using this method will have these three desirable properties.

3.2. Sharpness

The general procedure outlined in Section 2.1 guarantees that the coverage probability will never fall below $1-\alpha$. An important question, however, is whether or not this lower bound is ever achieved, i.e., are the intervals sharp? We first consider the binomial distribution.

In general, the binomial intervals will be conservative rather than sharp. This is illustrated in Figure 1 where, for $n=6$ and $\alpha=.10$, the coverage probabilities of the (sharp) one-sided interval and of the two-sided interval are displayed. It can be seen that if the nonconstant portions of the two one-sided curves overlap, the coverage probabilities at those points fall below $1 - (\alpha/2)$, the value for which the one-sided intervals are sharp. From Figure 1 we can also see when the two-sided intervals will achieve easily identified minimum coverage probabilities. There are two cases:

I. All lower bounds are strictly less than all upper bounds. In this case the two-sided intervals will achieve a minimum coverage probability of exactly $1 - (\alpha/2)$ at several values of θ.

II. The lower endpoints overlap upper endpoints in such a way that the combined coverage probability is α. In this case the two-sided intervals are sharp at $1-\alpha$.

Exploring when Case I holds is a relatively simple matter, since we only need examine when the greatest lower endpoint is less than the smallest upper endpoint, i.e, when does the inequality $\theta_L(n) < \theta_U(0)$ hold? Substituting into (2.15) and (2.16) shows that $\theta_L(n)$ and $\theta_U(0)$ are based on the same F-distribution, $F_{2,2n}$, and $\theta_L(n) < \theta_U(0)$ if and only if

$$F_{2,2n,\alpha/2} > n \quad , \tag{3.3}$$

which is equivalent to requiring

$$P(F_{2,2n} > n) \geq \alpha/2 \quad . \tag{3.4}$$

It is easy to see that for every n, there exists an $\alpha^*(n)$ such that the two-sided intervals attain minimum coverage probability $1 - (\alpha/2)$ for all $\alpha < \alpha^*(n)$. Similarly, given α there exists an $n^*(\alpha)$ such that the two-sided intervals attain minimum coverage probability $1 - (\alpha/2)$ for all $n < n^*(\alpha)$. Evaluating (3.4) is a relatively easy calculation, since the $F_{2,2n}$ density can be integrated exactly. In fact,

$$P(F_{2,2n} > n) = \int_0^\infty \frac{1}{(1 + \frac{1}{n}\, t)^{n+1}}\, dt = \frac{1}{2^n} \quad . \tag{3.5}$$

Thus, if $\alpha < 1/2^{n-1}$, using one-sided intervals at level $1 - (\alpha/2)$ will result in two-sided intervals that attain minimum coverage probability $1 - (\alpha/2)$. In general, however, unless n is very small, these α values are too small to be of practical use.

Next consider when Case II can occur. If, for some $x_1 < x_2$ we obtain $\theta_U(x_1) = \theta_L(x_2)$ then, at $\theta = \theta_U(x_1)$ we have

$$\begin{aligned} CP(\theta,\alpha) &= P_\theta(x_1 < X < x_2) \\ &= P_\theta(X < x_2) - P_\theta(X \le x_1) \\ &= \left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} \\ &= 1 - \alpha \quad , \end{aligned}$$

from the definitions of $\theta_U(x_1)$ and $\theta_L(x_2)$. Thus, if any lower endpoint ever equals an upper endpoint, the two-sided intervals are sharp at $1-\alpha$. Now, for every pair (x_1,x_2) with $x_1 < x_2$, there is a value of α for which $\theta_U(x_1) = \theta_L(x_2)$, making the intervals sharp for that value of α. Since there are $n(n-1)/2$ such pairs, there are exactly $n(n-1)/2$ values of α for which the intervals are sharp.

In general, for arbitrary $x_1 < x_2$, solving for α to obtain $\theta_U(x_1) = \theta_L(x_2)$ must be done numerically. However, for particular choices of x_1 and x_2 the equation is easily solved. Let $x_1 = x < n/2$, and let $x_2 = n-x$. Then, using (2.15) and (2.16), it can be seen that $\theta_U(x) = \theta_L(n-x)$ if and only if

$$F_{2(x+1),2(n-x),\alpha/2} = \frac{n-x}{x+1} \quad , \tag{3.6}$$

which is equivalent to the condition

$$P\left(F_{2(x+1),2(n-x)} > \frac{n-x}{x+1}\right) = \alpha/2 \quad . \tag{3.7}$$

Let $T = \Big((x+1)/(n-x)\Big)F_{2,(x+1),2(n-x)}$, and recall that $T/(1+T)$ is distributed as Beta$(x+1,n-x)$. Expressing (3.7) in terms of T, and then using the relationship between the beta and binomial given in (2.12), we can establish

$$P\left(F_{2(x+1),2(n-x)} > \frac{n-x}{x+1}\right) = P(Y \le x + 1) \quad , \tag{3.8}$$

where $Y \sim$ binomial $(n+1,\frac{1}{2})$. Thus, for this choice of x_1 and x_2, we see that the binomial confidence intervals are sharp for any α that satisfies

$$\alpha/2 = P(Y \le x+1) = (\tfrac{1}{2})^{n+1} \sum_{k=0}^{x+1} \binom{n+1}{k} . \tag{3.9}$$

Thus, (3.9) gives n+1 values of α, out of the possible $n(n-1)/2$ values, for which the binomial intervals are sharp. For example, if $n=10$, substituting $x=0,1$ into (3.9) yields $\alpha = .012,.065$, showing that these intervals would be sharp at these α-levels. While, in general, one would not choose α based on these considerations, the continuity (in α) of the interval endpoints at least holds some guarantee that the minimum coverage probability will be close to $1-\alpha$, whatever value is chosen for α.

The Poisson and negative binomial distributions are quite different cases from the binomial, the major reason being that they both have infinite sample spaces. In such cases, it turns out that one can always find sequences $\{x_i\}$ and $\{y_i\}$, $x_i < y_i$, such that $\lim\left(\theta_U(x_i)-\theta_L(y_i)\right) \to 0$, showing that the intervals are always sharp. While we offer no proof, the interested reader is referred to Casella and McCulloch (1984). Figures 2 and 3, which display the coverage probabilities of the Poisson and negative binomial intervals for $\alpha = .10$ are, in themselves, convincing evidence of this fact.

4. Conclusions

Explicit formulas for confidence intervals for the parameters of a binomial, Poisson, or negative binomial distribution are easily derived using the techniques presented here. Moreover, implementation for these distributions only requires χ^2 or F tables. The general technique outlined in Section 2.4 is, of course, applicable to many other discrete distributions, and although explicit formulas are not generally attainable, numerical approximations can be used.

The intervals constructed are nonrandomized intervals and, hence, will not possess such properties as uniformly most accurate unbiased. To obtain such intervals one must add random noise to the observed data, a requirement that seems, to say the least, dubious in practical applications.

The general, desirable, properties outlined in Section 3.1 are possessed by the intervals presented here. While these requirements may seem minimal, other techniques of constructing confidence intervals, such as inverting tests, may produce intervals lacking these properties. For example, Crow (1956) constructs binomial intervals by inversion of a test, and produces intervals which are not monotone in x.

Besides the properties of Section 3.1, the intervals for the Poisson and negative binomial parameters are sharp for all α. The intervals for the binomial parameter are usually conservative except for the few cases in which they are also sharp.

References

Blyth, C. R. and Still, H. A. (1983), "Binomial Confidence Intervals," *Journal of the American Statistical Association*, 78, 108-116.

Brownlee, K. A. (1960), *Statistical Theory and Methodology in Science and Engineering*, New York: Wiley.

Casella, G. and McCulloch, C. E. (1984), "A Useful Technique for Constructing Confidence Intervals in Discrete Distributions," Technical Report, Biometrics Unit, Cornell University, Ithaca, NY 14853.

Clemans, K. G. (1959), "Confidence Limits in the Case of the Geometric Distribution," *Biometrika*, 46, 260-264.

Clopper, C. J. and Pearson, E. S. (1934), "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, 26, 404-413.

Crow, E. L. (1956), "Confidence Intervals for a Proportion," *Biometrika*, 43, 423-435.

Garwood, F. (1936), "Fiducial Limits for the Poisson Distribution," *Biometrika*, 28, 437-441.

Ghosh, B. K. (1979), "Comparison of Some Approximate Confidence Intervals for the Binomial Parameter," *Journal of the American Statistical Association*, 74, 894-900.

Hald, A. (1952), *Statistical Theory with Engineering Applications*, New York: Wiley.

Johnson, N. L. and Kotz, S. (1969), *Discrete Distributions*, Boston: Houghton-Mifflin.

Kendall, M. G. and Stuart, A. (1979), *The Advanced Theory of Statistics*, Vol. 2, London: Charles Griffin.

Miettinen, O. S. (1970), "Estimation of Relative Risk from Individually Matched Series," *Biometrics*, 26, 75-86.

Miller, R. G. Jr. (1977), *Simultaneous Statistical Inference*, 2nd Ed., New York: Springer-Verlag.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974), *Introduction to the Theory of Statistics*, New York: McGraw-Hill.

Neter, J., Wasserman, W., and Whitmore, G. A. (1982), *Applied Statistics*, Boston: Allyn and Bacon.

Snedecor, G. W. and Cochran, W. G. (1980), *Statistical Methods*, 2nd Ed., Ames: Iowa State Press.

Stevens, W. L. (1950), "Fiducial Limits of the Parameter of a Discontinuous Distribution," *Biometrika*, 37, 117-129.

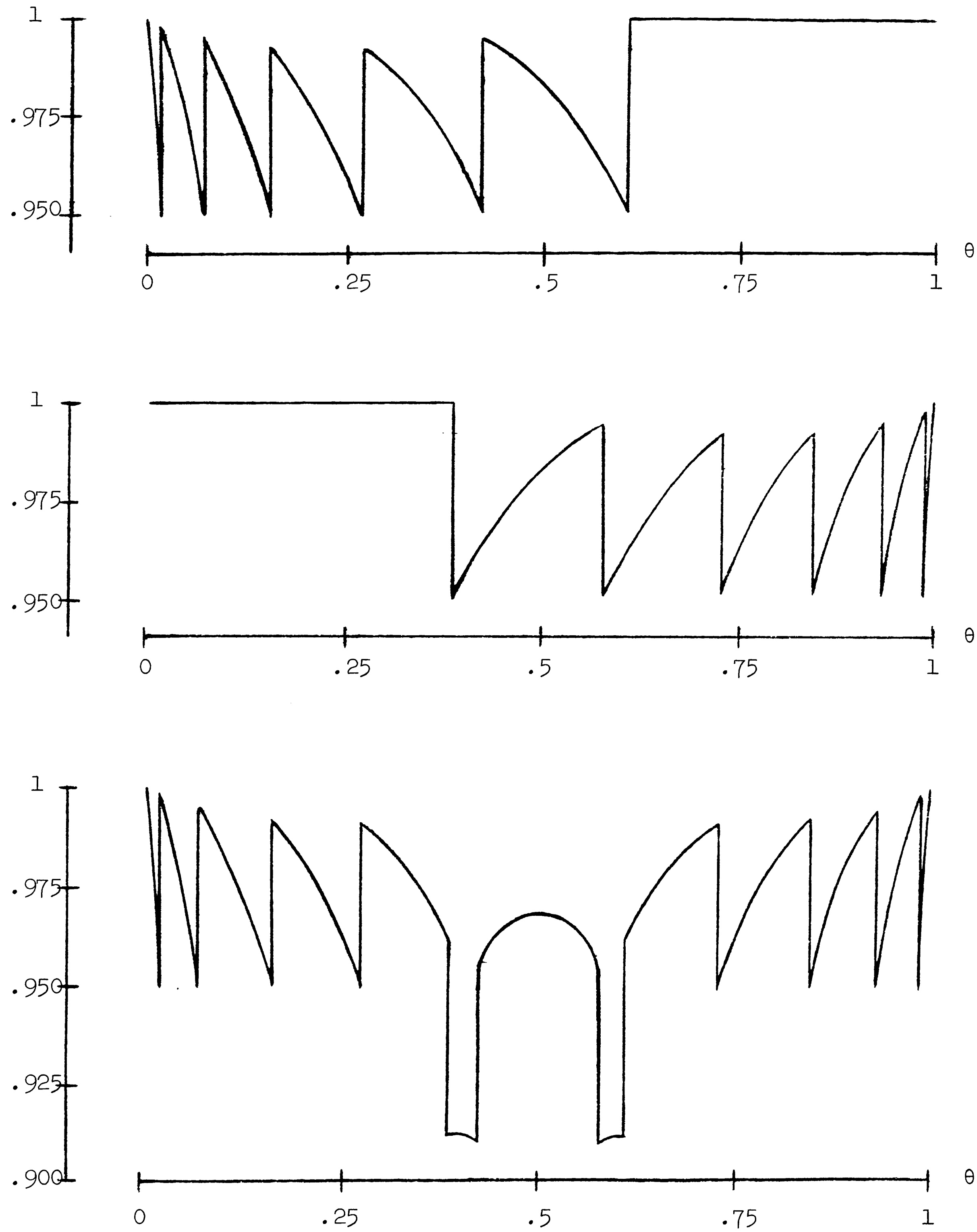


Figure 1. Coverage probabilities for lower, upper and two-sided binomial intervals, $\alpha = .1$, $n = 6$.

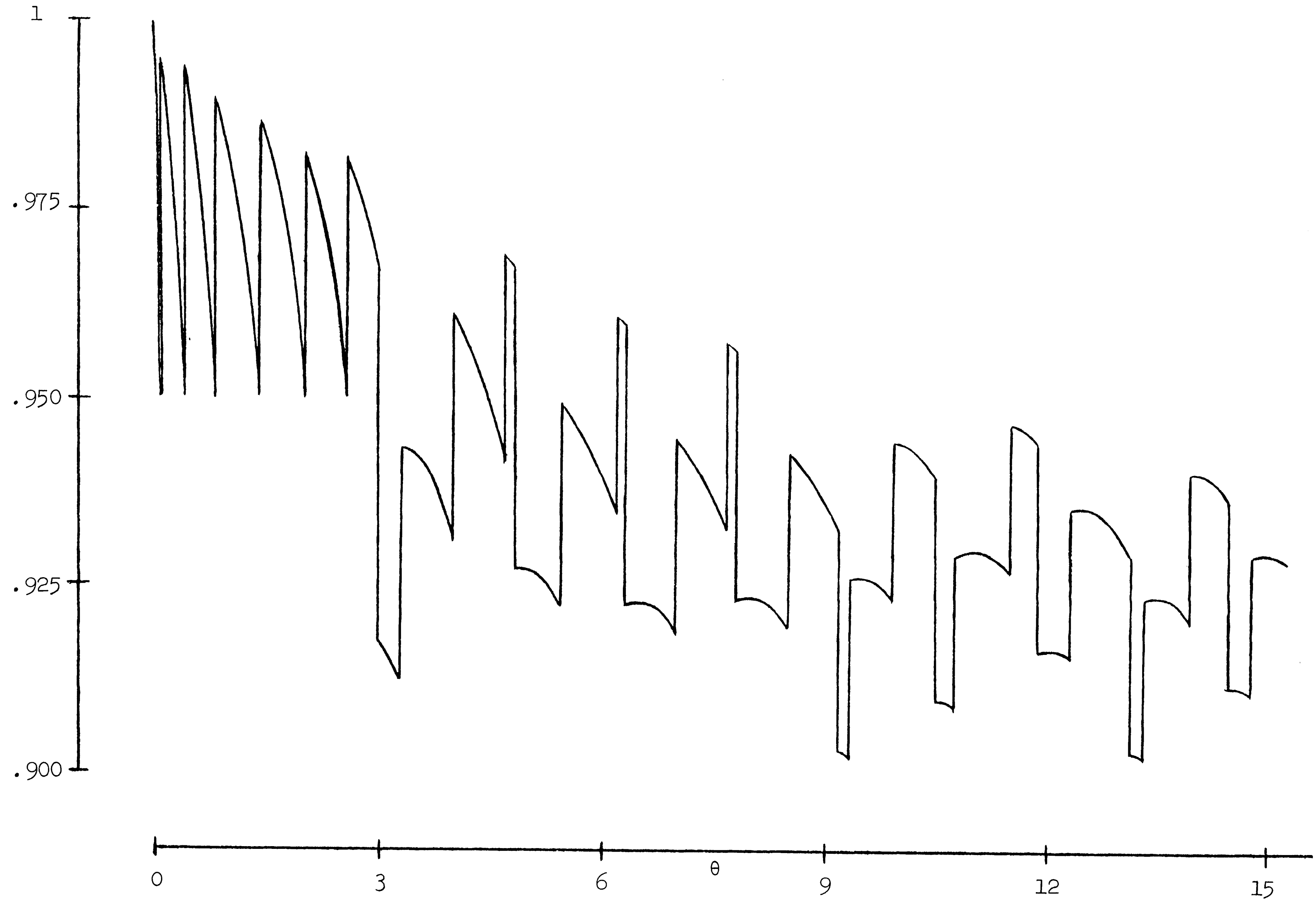
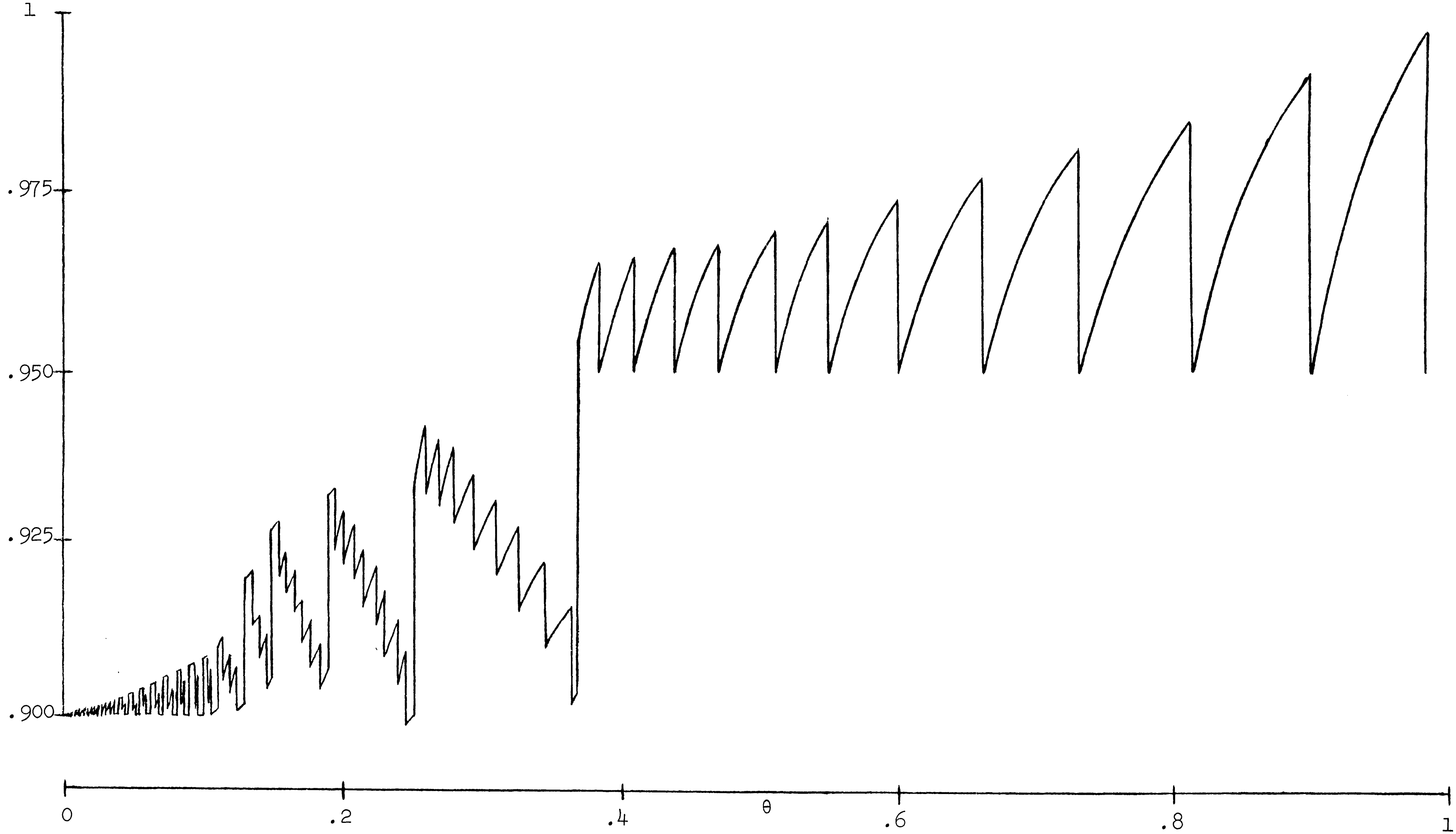


Figure 2. Coverage probabilities for Poisson intervals, $\alpha = .1$.



<u>Figure 3</u>. Coverage probabilities for negative binomial intervals, $\alpha = .1$, $r = 3$.